

КОРПУС ТЭКСТАЎ У ФІЛАЛАГІЧНЫМ ДЫСКУРСЕ

Дзейнасць па арганізацыі корпуса тэкстаў можа быць дастаткова разнастайнай. Крытэрыі адбору тэкстаў для свайго корпуса стваральнік задае адпаведна скіраванасці сваёй практычнай ці навуковай дзейнасці: парог адлюстравання будзе вызначацца патрабаваннем прадставіць у корпусе запланаваны моўны аб'ект. Нават аб'яднанне ў адным файле адабраных па пэўных крытэрыях тэкстаў адбываецца, як правіла, з улікам папярэдне вызначанай мэты, што немагчыма без усведамлення металінгвістычных умоў стварэння і функцыянавання абранага матэрыялу. Разам з тым, пачынаючы з Браўнаўскага корпуса — амаль адразу выкарыстанага для філасофскіх даследаванняў, — практычная вартасць створанага корпуснага рэсурса, як правіла, перавышае запланаваныя вынікі [3].

Тым не менш, да гэтага часу недаацэнена магчымасць выкарыстання корпусных рэсурсаў для дыскурсіўных практык, кантэнт-аналізу, экспертных сістэм і многіх іншых прыкладных даследаванняў. Інновацыйны патэнцыял корпусных даследаванняў з цягам часу будзе толькі ўзрастаць.

Паўвекавы тэрмін корпуснай дзейнасці абумовіў цікавую асаблівасць адпаведнай дыскурсіўнай практыкі: узнаўленне раней створаных рэсурсаў. Корпусы, якія працуюць на базе маральна састарэлага праграмнага забеспячэння, нярэдка перарабляюцца пад новыя тэхналагічныя магчымасці. Напрыклад, стварэнне «Цюбінгенскіх корпусаў рускіх тэкстаў» на базе «Упсальскага корпуса сучасных рускіх тэкстаў», сведчаць аб дастаткова высокай філалагічнай вартасці тэкстаў, як самадастатковай і ўнікальнай сукупнасці [2].

Дыскурсіўны аспект корпусных распрацовак мае важнае значэнне, аказваючы сур'ёзны ўплыў на даследчы працэс у корпусным рэчышчы. Паводле крытэрыяў адбору тэкстаў і рэпрэзентатыўнасці корпусы прынята дзяліць на два класы. У першы клас уваходзяць універсальныя корпусы тэкстаў, мэта якіх — аб'ектыўна адлюстроўваць шырокую плынь маўленчай дзейнасці. Толькі суцэльны, без суб'ектыўнага адбору корпус, напрыклад, нацыянальны корпус, дазваляе атрымаць поўную, сістэмную карціну стану мовы. Прычым, на практыцы корпусныя даследаванні могуць насіць лакальны па пэўных паказчыках характар. Метадалогія пабудовы корпусаў першага тыпу так ці інакш грунтуецца на прынцыпе дэдукцыі — ад агульнага (аб'ектыўнай маўленчай практыкі) да канкрэтных тэкстаў, якія адлюстроўваюць гэта агульнае.

Здольнасць корпуса тэкстаў адэкватна адлюстроўваць уласцівасці праблемнай вобласці, рэлевантныя для вызначанага тыпу лінгвістычнага даследавання, вызначаецца яго рэпрэзентатыўнасцю. Ва ўніверсальных корпусах высокая дакладнасць вынікаў забяспечваецца высокімі колькаснымі паказчыкамі. Неабходнасць улічваць пэўную прапарцыянальнасць і структураванасць тэкстаў абумоўлена частотнасцю той ці іншай з'явы ў праблемнай вобласці. Гэта патрабаванне арыентуе «збіральнікаў» корпусаў тэкстаў на папярэднюю спецыялізацыю ствараемага рэсурса па ўзроўневым прынцыпе: фанетычную, марфалагічную, сінтаксічную, лексічную, семантычную, дыскурсіўную і г. д.,

што, дарэчы, не выключае ўніверсальную арыентацыю моўнага матэрыялу згодна вызначаным мэтам.

У другі клас уваходзяць ілюстратыўныя корпусы, пабудаваныя для адлюстравання пэўнай з’явы. У гэтым выпадку прыходзіцца задаволіцца тымі ці іншымі выбаркамі з агульнай сукупнасці моўных дадзеных і ў рэчышчы індукцыі рухацца да абагульненняў. Але гэтыя выбаркі павінны адпавядаць крытэрыям якаснай прадстаўнічасці выбаркі для лінгвістычнага даследавання. Менавіта разгляд праблемнай вобласці з улікам дыскурсіўнага аспекту дазваляе паспяхова вырашаць многія лінгвістычныя задачы на аснове кампактнага прадстаўнічага матэрыялу. Па гэтай прычыне метадалогія пабудовы корпусаў другога класа павінна мець высокую ступень карэктнасці адлюстравання асобнага лінгвістычнага феномена.

Сучасны трэнд лінгвістыкі, эмпірычныя даследаванні, абумоўліваюць цікавасць да збору і сістэматызацыі вялізных корпусных масіваў, у якіх парадак лічбаў словаўжыванняў — сотні мільёнаў. Напрыклад у «Нацыянальным корпусе рускай мовы» — прыкладна паўмільярда лексічных адзінак [1]. Найлепей такой тэндэнцыі адпавядаюць корпусы першага тыпу — універсальныя, нацыянальныя. Але і спіс корпусаў, створаных замежнымі лінгвістамі для пэўнай прагматычнай мэты, проста бясконцы, прычым большая частка з іх, што характэрна, пазней выкарыстоўваецца для мэтаў, больш шырокіх за першапачатковыя.

Поўныя ці рэпрэзентатыўныя моўныя корпусы і корпусы, якія ствараюцца для прыватных лінгвістычных задач, як правіла, маюць сінкрэтычныя мэты. З гэтага лагічна вынікае відавочная і важная выснова аб эфектыўным выкарыстанні ўжо акумуляваных рэсурсаў. Нават калі для пэўнай мовы (напрыклад, беларускай) яшчэ не створаны ўніверсальныя ці, нават, нацыянальны корпус, то прыватныя даследчыя корпусы ёсць сэнс рабіць агульнадаступнымі: найлепшы для гэтага сродак сёння — Інтэрнэт. Даступнасць корпусных дадзеных у Інтэрнэце будзе садзейнічаць іх верыфікацыі.

Экстралінгвістычныя асаблівасці корпуснай дзейнасці экстрапаліруюцца і на самі корпусы. Рэпрэзентатыўнасць любога корпуса, у прынцыпе, не абмяжоўваецца ўласналінгвістычнымі параметрамі. Так, у кожным канкрэтным выпадку можа ўзнікнуць патрэба ўлічыць стылістычны, часавы, аўтарскі і іншыя элементы тэкставага масіву праблемнай вобласці. Наяўнасць экстралінгвістычнай разметкі тэкстаў корпуса — своеасаблівай абалонкі корпусных дадзеных — дазваляе падпарадкаваць працу даследчыкаў вопыту і кампетэнцыі яго складальнікаў.

Успрыманне корпуснай сукупнасці дадзеных ў кантэксце, з улікам дыскурсіўных якасцей, стварае аснову для сістэмнага даследавання маўленчай дзейнасці. І такая парадыматычная абумоўленасць корпусных рэсурсаў ператварае сукупнасць тэкстаў у шматузроўневы культурны феномен.

Літаратура

1. Национальный корпус русского языка [Электронны рэсурс]. — Рэжым доступу: <http://www.ruscorpora.ru>. — Дата доступу: 29.08.2013.
2. Тюбингенские корпусы русских текстов [Электронны рэсурс]. — Рэжым доступу: <http://www.sfb441.uni-tuebingen.de/b1/rus/korpora.html>. — Дата доступу: 29.08.2013.
3. Brown Corpus Manual [Электронны рэсурс]. — Рэжым доступу: <http://khnt.aksis.uib.no/icame/manuals/brown>. — Дата доступу: 29.08.2013.